Taylor & Francis
Taylor & Francis Group

International Journal of Green Energy

ISSN: 1543-5075 (Print) 1543-5083 (Online) Journal homepage: http://www.tandfonline.com/loi/ljge20

Optimizing Mixture Properties of Biodiesel Production Using Genetic Algorithm-Based Evolutionary Support Vector Machine

Min-Yuan Cheng, Doddy Prayogo, Yi-Hsu Ju, Yu-Wei Wu & Sylviana Sutanto

Optimizing Mixture Properties of Biodiesel Production Using Genetic Algorithm-Based Evolutionary Support Vector Machine

Min-Yuan Cheng[1], Doddy Prayogo[2*], Yi-Hsu Ju[3], Yu-Wei Wu[4]

Professor, Dept. of Civil and Construction Engineering, National Taiwan University of Science and Technology, Taiwan.

#43, Sec. 4, Keelung Rd., Taipei, Taiwan, R. O.C. 106, Dept. of Civil and Construction Engineering

Dept. of Civil and Construction Engineering, National Taiwan University of Science and Technology, Taiwan.

Dept. of Civil Engineering, Petra Christian University, Indonesia

Jalan Siwalankerto 121-131, Surabaya 60236, Indonesia, Dept. of Civil Engineering., Taipei, Taiwan, R. O.C. 106, Dept. of Civil and Construction Engineering

Jalan Siwalankerto 121-131, Surabaya 60236, Indonesia, Dept. of Civil Engineering

Dept. of Chemical Engineering, National Taiwan University of Science and Technology, Taiwan.

#43, Sec. 4, Keelung Rd., Taipei, Taiwan, R. O.C. 106, Dept. of

Chemical Engineering

Postdoctoral researcher, Dept. of Civil and Construction Engineering, National Taiwan University of Science and Technology, Taiwan.

#43, Sec. 4, Keelung Rd., Taipei, Taiwan, R. O.C. 106, Dept. of

Civil and Construction Engineering

Dept. of Chemical Engineering, National Taiwan University of Science and Technology, Taiwan.

#43, Sec. 4, Keelung Rd., Taipei, Taiwan, R. O.C. 106, Dept. of

Chemical Engineering

*Corresponding author yhju@mail.ntust.edu.tw

Abstract

Nowadays, biodiesel is used as one of the alternative renewable energy due to the increasing energy demand. However, optimum production of biodiesel still requires a huge number of expensive and time-consuming laboratory tests. To address the problem, this research develops a novel Genetic Algorithm-based Evolutionary Support Vector Machine (GA-ESIM). The GA-ESIM is an Artificial Intelligence (AI)-based tool that combines K-means Chaotic Genetic Algorithm (KCGA) and Evolutionary Support Vector Machine Inference Model (ESIM). The

ESIM is utilized as a supervised learning technique to establish a highly accurate prediction model between the input-output of biodiesel mixture properties; and the KCGA is used to perform the simulation to obtain the optimum mixture properties based on the prediction model. A real biodiesel experimental data is provided to validate the GA-ESIM performance. Our simulation results demonstrate that the GA-ESIM establishes a prediction model with better accuracy than other AI-based tool and thus obtains the mixture properties with the biodiesel yield of 99.9%, higher than the best experimental data record, 97.4%.

Keywords: Biodiesel production, Rice bran, In-situ process, Genetic Algorithm, Evolutionary Support Vector Machine

1. Introduction

Nowadays, the world encounters energy demand problems due to increasing population and global economic development. Therefore, the remarkable global energy demand triggers the over consumption of fossil fuel. This phenomenon yields the following main issues: (1) excessive greenhouse emissions and air pollutions, (2) global warming and climate change, and (3) fossil fuel depletion (Puig-Arnavat, Bruno, and Coronas 2010; Ching-Piao et al. 2012; Noam 2010; Baños et al. 2011). Facing these serious problems, the use of renewable energy as an alternative energy supply must be developed. Biodiesel has been recognized for its ability in yielding energy with less environmental impacts than fossil fuel (Demirbas 2007; Gaurav, Srivastava, and Singh 2012; Al-Mulali 2014). Moreover, with recent inflation of current fossil fuel prices, uncertainties regarding its future availability, and the need for environment friendly fuels, there is an increased attention in utilizing biodiesel as alternate fuel (Hoekman et al. 2011). Since biodiesel can be produced by in-expensive raw material such as rice bran, the agricultural waste in most rice producing countries (Shiu et al. 2010; Wakil et al. 2014; Chuah et al. 2006), production cost might not become the major threat for its commercialization.

A number of studies present that the production cost of biodiesel might be decreased even more by applying an in-situ process (Ozgul-Yucel and Turkay 2002; Yustianingsih, Zullaikah, and Ju 2009). In this process, additional mixtures such as alcohol and catalyst are usually needed. Consequently, expensive and time-consuming laboratory test are required to optimize of those properties. Several researchers need to provide a number of sample tests to conduct experiments involving different mixtures to indentify the influencing variables that affecting the biodiesel yield (Freedman, Pryde, and Mounts 1984; Dorado et al. 2004). It was found that the relationship among production of biodiesel and its corresponding mixture properties become more complex due to the large number of components. Therefore, there is a need to implement a model that is able to predict the complex relationship accurately.

Recently, the adoption of Artificial Intelligence (AI) approaches in biodiesel fields has attracted the attention from the researchers. Hybrid Artificial Neural Network - Genetic Algorithm (ANN-GA) is example of AI approaches that have been successfully applied to solve complex problems in biodiesel fields (Rajendra, Jena, and Raheman 2009). However, both ANN and GA have their main disadvantages. For example, improper parameters setting of ANN can decrease the prediction accuracy. On the other hand, premature convergence and trapped in local optima problems are often faced by GA. This has thus encouraged many researchers to conduct a number of studies in improving the quality of AI techniques.

K-means and Chaos Genetic Algorithm (KCGA) (Cheng and Huang 2010) and Evolutionary Support Vector Machine Inference Model (ESIM) (Cheng and Wu 2009) are examples of AI approaches that have been applied in several engineering fields. KCGA is a powerful hybrid optimization algorithm that minimizes some shortcomings of traditional GA while ESIM improves the performance of Support Vector Machine (SVM) in dealing with complex input-output relationship by self-tuning the SVM parameters automatically. The success of KCGA and ESIM for improving the performance of GA and SVM opened the possibilities of establishing more advance hybrid algorithm.

This paper introduces a novel hybrid Genetic Algorithm-based Evolutionary Support Vector Machine (GA-ESIM) which combines KCGA and ESIM for optimizing biodiesel mixture properties. ESIM is employed for building accurate prediction model whilst KCGA is performed to search the optimum mixture properties. Thus, Integrating KCGA and ESIM can offer strong potential to generate a robust method for searching the optimum mixture properties of biodiesel.

2. Artificial Intelligence for biodiesel

In recent years, various studies have been conducted in the field of in-situ biodiesel production (Ozgul-Yucel and Turkay 2002; Yustianingsih, Zullaikah, and Ju 2009; Shiu et al. 2010). Those studies have shown that the in-situ biodiesel production has successfully been tested in several circumstances. However, although these approaches reduce further production cost of biodiesel, a huge number of laboratory tests are still required in order to get the optimum production of biodiesel. Therefore, to solve this situation, a model that can predict and search the optimized mixture properties without implementing expensive and time-consuming tests must be developed.

Many studies have proposed Artificial Intelligence (AI) approaches, such as Artificial Neural Network (ANN) or Support Vector Machine (SVM), as alternative methods to solve complex and ill-defined problems. Those approaches have been widely proven its usefulness in handling complex input-output relationship. In recent years, the use of ANN has been developed in solving many problems in the field of renewable energy, such as solar radiation, wind-speed prediction, photovoltaic systems, and biomass gasification (Kalogirou 2001). As for other renewable energy such as biodiesel, there are significant numbers of studies that have been conducted.

Ramadhas et al. (2006) proposed an ANN to predict the cetane number (CN) of biodiesel. It was found that the ANN models developed could be used reliably for the prediction of biodiesel CN (Ramadhas et al. 2006). Yuste and Dorado (2005) demonstrated the superiority of ANN in predicting biodiesel yield through the transesterification of used frying olive oil. It was reported that the results were similar to those obtained with the help of the empirical tests required in a laboratory, thus, indicating ANN has proven to be capable of modeling the production of biodiesel from waste olive oil, which is a process with a high nonlinear behavior (Yuste and Dorado 2005).

Instead of using solely ANN for predicting the input-output relationship of biodiesel production, a number of studies present an ANN coupled with optimization technique for different purpose. Rajendra et al. (2009) use hybrid ANN-GA to optimize the input parameters of biodiesel production (Rajendra, Jena, and Raheman 2009). The input parameters for the ANN to generalize the pretreatment process were initial acid value of vegetable oil (IAV), methanol-to-oil ratio (M), catalyst concentration (C), and reaction time (T); and the output parameter was final acid value (FAV) of oil. After ANN was performed to develop the input-output relationship, initial random population was created by GA and then given to the developed ANN for predicting the FAVs. Then, the usual genetic operators (selection, crossover, and mutation) were performed until termination criterion was achieved. As a result, the proposed hybrid model was able to find the optimum mixtures and verified by laboratory experiment.

Despite being widely used, both ANN and GA have their main shortcomings. To this end, more advance hybridization of prediction and optimization algorithms are still needed to optimize mixtures of biodiesel.

3. Genetic Algorithm-based Evolutionary Support Vector Machine (GA-ESIM)

This section explains the proposed model in detail. GA-ESIM combines KCGA and ESIM to build a model that can search an optimal solution from the dataset which contains the complex input-output relationship. The proposed model employs KCGA as the optimizer and ESIM for mapping the relationship. The hybridization of those AI methods is expected to produce a robust model for optimizing mixture properties. The following section first briefly introduces two main core AI methods, namely KCGA and ESIM, and explains the proposed GA-ESIM in detail.

3.1. K-means and Chaos Genetic Algorithm (KCGA)

KCGA is a powerful hybrid algorithm proposed by Cheng and Huang (2009) which integrates k-means and chaos attributes based on GA (Cheng and Huang 2009). In this hybrid algorithm, k-means plays a critical role in convergence of GA whereas chaos algorithm can keep GA population diversity and avoid from premature convergence. Initial individuals of KCGA were generated using chaos algorithm to diversify their positions between the lower and upper bound of the domain value. The individuals were evaluated and ranked. Then, crossover and mutation were performed. The chaos operator diversified individuals by using logistic map and eventually generates chaotic spread-spectrum and unpredictable irregular motions of each individual. K-means clustering groups the individuals into *k* mutually exclusive clusters and locates the centroid of each cluster. Thus, location information of each cluster centroid would be treated as candidate individuals for the next generation. A competing procedure was employed to eliminate lower fitness value individuals, and reserved the others to create formal population for KCGA iteration.

KCGA obtain more accuracy by enhancing the diversity of GA using chaos mapping. KCGA also further extracts clustering rules for achieving a potential trend of evolution. As a result, KCGA can effectively solve some drawbacks of traditional GA, such as long running time and being trapped in local optima.

3.2. Evolutionary Support Vector Machine Inference Model (ESIM)

Support vector machine (SVM) is a machine-learning algorithm combining Vapnik-Chervonenkis Dimension of statistics with Structure Risk Minimization Theory that was widely adopted after being proposed by Vapnik (Vapnik 1995). SVM classifies data using different class labels by determining a set of support vectors, which are members of the set of training inputs that outline a hyper plane in a feature space. Furthermore, it provides a generic mechanism that fits the hyper plane surface to the training data using a kernel function (Huang and Wang 2006).

However, SVM presents users the problem of tuning optimal kernel parameters for users. The proper parameter can increase SVM prediction accuracy greatly, therefore, SVM parameters must be optimized simultaneously. The parameters includes penalty parameter C and radial basis function (RBF) kernel parameters γ. Fast messy genetic algorithms were developed by Goldberg et al. in 1993 (Goldberg et al. 1993). Unlike the well-known simple genetic algorithm (sGA), which uses fixed length strings to represent possible solutions, fmGA applies messy chromosomes to form strings of various lengths. Its ability to identify optimal solutions efficiently for large-scale permutation problems gives fmGA the potential to generate parameter C and γ of SVM simultaneously.

Taking the benefits of the two AI approaches, a hybrid SVM-fmGA, namely ESIM, was proposed by Cheng and Wu (Cheng and Wu 2009). In ESIM, the SVM is employed primarily to address the learning and curve fitting while fmGA addresses optimization. Furthermore, this model was developed to achieve the fittest C and γ parameters with minimal prediction error.

3.3. Genetic Algorithm-based Evolutionary Support Vector Machine Algorithm

The implementation GA-ESIM is a two-step approach, which involves the initialization and optimization phase. In initialization phase, ESIM is utilized to map out the input-output relationship of the data involved. As a result, a prediction model is established, including the optimal C and γ parameters. This prediction model can predict the percentage of FAME yield given the required input variables. The output prediction model will be carried out and shall be utilized in the next phase for predicting the optimum mixture component which will produce the highest percentage of FAME yield.

The second phase of GA-ESIM is the optimization phase. Once the input-output relationship of the dataset is established, then, KCGA is employed as a search and optimization algorithm. The role of KCGA is to find the optimum mixture component which yields the fittest value. The individuals of population represent the feasible solutions. These individuals are generated and evaluated using the established ESIM prediction model. The population was initialized randomly. Each individual represents a potential solution to a problem. Then, all KCGA parameters were set, including population size, number of generation, stopping criteria, crossover rate, mutation rate, and number of k-means cluster. The population will be evaluated to obtain the fitness value of each individual. In a number of complex problems, the objective function cannot be expressed in mathematical equation. In this situation, ESIM plays a role to map the each individual and its fitness value using the optimal C and γ. The population of solution undergoes selection, crossover, mutation, chaos operator and k-means clustering until the fittest solution, that is the mixture with the highest FAME yield, is obtained.

Finally, the complete GA-ESIM algorithm flowchart is shown in Fig. 1.

4. Result and Discussion

This section demonstrates the performance of the proposed model. This study uses GA-ESIM to conduct a biodiesel production simulation for searching the optimum mixture properties. This section is organized to describe the explanation of model application in detail.

4.1. Input Data

This paper uses the original research data of biodiesel production from rice bran which were conducted by Shiu et al (Shiu et al. 2010). This experiment has successfully produced biodiesel from rice bran by a two-step in-situ reaction (acid-catalyzed followed by base-catalyzed). The highest FAME yield of 97.4% was obtained after evaluating laboratory tests, which were conducted within various conditions.

Table 1 shows examples of several data from biodiesel experiment. The collected database contains total 79 records, includes 38 records of one-step in-situ process and 41 records of two-step in-situ process. One-step in-situ process involves the only in-situ acid- catalyzed esterification. Meanwhile, the other method is two-step process in which in-situ acid-esterification was directly followed by in-situ transesterification without a separation step in between.

Since the one-step process only conducts the in-situ esterification, the value of both NaOH content and second step reaction attributes are certainly 0. Moreover, it was found that the same mixture solutions in the database were tested several times. For the example, three first cases which consist of the same mixture solutions; one-step process, initial FFA content of 3%, methanol to rice bran ratio of 2.5 mL/g, sulfuric acid to rice bran mass ratio of 27.6%, 60 minutes first step reaction; were reacted and produced slightly different output results.

The database covers 8 attributes as shown in Table 2, 7 of which are input factors, and the output factor is FAME yield. In this paper, the input and output variables were normalized between 0 and 1 in order to avoid the numerical difficulties or condition where attributes with greater ranges dominating those with smaller ranges (Hsu, Chang, and Lin 2003). The function used to normalize the data is shown in Eq. (1).

$$x_n^{norm} = \frac{X_n - X_n^{min}}{X_n^{max} - X_n^{min}} \quad (1)$$

where x_n^{norm}= normalized data of attribute *n*, x_n = initial data of attribute *n*, x_n^{max}= upper bound data of attribute *n*, x_n^{min} = lower bound data of attribute *n*.

<*Insert* Table 2 *here*>

4.2. Cross validation to verify ESIM training

Since partitioning data randomly into training and testing dataset presents certain disadvantages, cross validation techniques are often used to solve those shortcomings. One of the disadvantages was whenever prediction error is excessive, the testing data must be re-sampled until error conditions are satisfied. Thus, only the best model can be used to predict unknown cases. This may cause bias of actual model error. Thus, researchers often use *k*-fold cross validation to minimize the bias. Kohavi showed that 10 folds were optimal to obtain the minimal time needed to perform the test with acceptable bias associated with the validation process (Kohavi 1995).

The dataset introduced in Section 4.1 were subjected to 10-fold cross validation in order to ensure that all the dataset were applied in both the training and testing phase. The procedure was performed in the following four steps:

1. Randomize the data;

2. Divide the data into 10 folds which have the equal amount and number each fold from 1 to 10;

3. Generate single set by assigning one fold as the testing case and the other nine folds as training cases. When fold 1 was considered the testing dataset, folds 2 through 10 were training dataset. When fold 2 was the testing dataset, the other nine folds were treated as training dataset, and so on.

4. Repeat step 3 until generate 10 folds consists of 10 sets of different training and testing dataset.

Total 79 records of biodiesel experiments were divided into 10 folds. The first nine folds have a total of 71 training records and 8 testing records, while the last fold has 72 training records and 7 testing records. ESIM were employed to run the total 10 folds. The Root Mean Square Error (RMSE) was employed to evaluate an error measure of each fold. The formulation of RMSE is expressed in the following equation:

$$\text{RMSE} = \sqrt{\frac{\sum (y' - y)^2}{n}} \quad (2)$$

where *y* = actual value; *y'* = predicted value; and *n* = number of data samples.

4.3. ESIM prediction model

This section presents the prediction results of ESIM training and testing. In addition, well-known prediction tool, namely SVM, was also employed for comparison purposes. The *C* and γ parameter for SVM were set to 1 and 0.14286 suggested by Hsu et al. (Hsu, Chang, and Lin 2003). As for ESIM, the parameters were automatically optimized to achieve the minimum prediction error. Table 3 shows RMSE results of the ESIM compared against SVM.

ESIM produces less error than SVM. This indicates that the new model achieves better performance for predicting FAME yield percentage than SVM. For training results, the minimum, average and maximum RMSE of ESIM obtained were 6.98, 9.54, and 12.20 respectively, whereas for testing results, 4.80, 6.49, and 8.55 were obtained. The best prediction result for biodiesel production was found in fold 3, which has the lowest testing RMSE of 4.80. Table 4 and Table 5 show the training and testing result of the best fold in detail. ESIM training and testing results of fold 3 were plotted in Fig. 2.

Finally, ESIM C and γ parameters in the best fold, 6 and 0.2647 respectively, were identified as the optimal tuning parameters. In the last step, the optimal C and γ parameters will be carried out to GA-ESIM step as the optimal prediction model. With this optimal prediction model, ESIM is ready for determining the new input-output relationship during the optimization process.

4.4. The optimization process of GA-ESIM

With the ability for mapping out the input-output relationship, ESIM has two roles in this proposed model. Firstly, ESIM used the mapping ability to train the biodiesel data to create prediction model before the optimization started. Secondly, during the optimization process conducted by KCGA, ESIM used the prediction model to map out the relationship between mixture properties as the input and FAME yield as the output of biodiesel data. Finally, the optimization process continued until the optimal mixture solution with the best FAME yield is found.

KCGA is employed due to the ability to improve the performance of traditional GA with the help of two additional operators, namely chaos operator and k-means clustering. These operators ensure the diversity of the population during the optimization process and speed up the convergence to global optima. The role of KCGA is to find the optimum mixture component which yields the highest FAME. Feasible solutions are generated and evaluated using the established ESIM prediction model. Solutions with the higher FAME yield are regarded as fitter solutions. In KCGA, the population of solution undergoes selection, crossover, mutation, chaos operator and k-means clustering until the fittest solution, that is the mixture with the highest FAME yield, is obtained.

In this study, parameters of KCGA were set as follows: number of generation = 100, population size = 40, crossover rate = 0.9, mutation rate = 0.05, and number of k-means cluster = 3. Reaching maximum number of generations was applied to stop the computational work. Each individual in the population of KCGA represented the random biodiesel mixture components. The upper and lower bound of the input variables is adopted from the experiment.

Then, optimal C and γ parameters obtained from Section 4.3 (C = 6 and γ = 0.2647) were used to map each individual to its fitness value. KCGA operators, including selection, crossover, mutation, chaos, and k-means, were employed to search the best individual. The process was repeated until the stopping criterion was met. The convergence graphic is shown in Fig. 3. Finally, the result of the optimum mixture obtained was listed and compared to the original laboratory research in Table 6.

The maximum FAME yield of 99.9% was obtained by GA-ESIM after 100 generations. The result shows the better FAME yield value comparing with the maximum value of 97.4% of the FAME yield produced in the real laboratory tests. This implies that performing test using the mixture properties as given in Table 5, will achieve the increasing FAME yield of biodiesel production by 2.1% with the RMSE of 4.80%.

4.5. Integrating GA-ESIM with Graphic User Interface (GUI) of MATLAB

A graphical user interface (GUI) is a graphical display that enables users to perform interactive tasks. This study developed a system that integrates GA-ESIM with GUI to provide users easier and more effective interaction. To demonstrate the performance of the system, an example is provided as follows: lower and upper bound, KCGA and ESIM parameters were set as the same setting as in Section 4.4.

There are two main panels in the interface, input data and output data panel. In input data panel, users can input: (1) lower bound and upper bound of each component; (2) KCGA parameter such as the number of generation, crossover rate, mutation rate, population size and number of clusters; (3) ESIM parameter C and γ. The output data panel includes: (1) optimum mixture contents; (2) result of predicted FAME yield and computational time. The mixture of biodiesel production is presented on the output panel of the proposed system as shown in Fig. 4.

5. Conclusions

This research developed the GA-ESIM for optimizing biodiesel mixture properties by fusing KCGA together with ESIM. ESIM primarily achieved the concurrently C and γ parameters to build the accurate prediction model verified by cross validation. KCGA can conduct simulations of trial mixes and searches the optimum solution in short time without being trapped in local optima. In this case study, individuals of KCGA represent compositions of biodiesel mixture properties while the fitness value represent FAME yield. During the optimization process performed by KCGA, ESIM determines the complex relationship between each individual and its fitness value.

The comparison between ESIM and SVM showed the superiority of ESIM as a prediction tool for biodiesel production. Thus, the incorporation between ESIM together with KCGA, which were proven to outperform traditional GA, shows the strong potential of GA-ESIM as a robust model for optimizing biodiesel mixture properties. In addition, this model does not rule out the possible application in other academic and engineering fields.

References

Al-Mulali, Usama. 2014. The Impact of Biofuel Energy Consumption on GDP Growth, Co2 Emission, Agricultural Crop Prices, and Agricultural Production. *International Journal of Green Energy* 12 (11):1100-1106.

Baños, R., F. Manzano-Agugliaro, F. G. Montoya, C. Gil, A. Alcayde, and J. Gómez. 2011. Optimization methods applied to renewable and sustainable energy: A review. *Renewable and Sustainable Energy Reviews* 15 (4):1753-1766.

Cheng, Min-Yuan, and Kuo-Yu Huang. 2009. K-means clustering and Chaos Genetic Algorithm for Nonlinear Optimization. Paper read at The 26th International Symposium on Automation and Robotics in Construction (ISARC 2009).

Repeated Author. 2010. Genetic algorithm-based chaos clustering approach for nonlinear optimization. *Journal of Marine Science and Technology* 18 (3):435-441.

Cheng, Min-Yuan, and Yu-Wei Wu. 2009. Evolutionary support vector machine inference system for construction management. *Automation in Construction* 18 (5):597-604.

Ching-Piao, Tsai, Hwang Ching-Her, Hwa Chien, and Cheng Hao-Yuan. 2012. Study on the wave climate variation to the renewable wave energy assessment. *Renewable Energy* 38 (1):50-61.

Chuah, T. G., A. G. K. Wan Azlina, Y. Robiah, and R. Omar. 2006. Biomass as the Renewable Energy Sources in Malaysia: An Overview. *International Journal of Green Energy* 3 (3):323-346.

Demirbas, Ayhan. 2007. Recent Developments in Biodiesel Fuels. *International Journal of Green Energy* 4 (1):15-26.

Dorado, M. P., E. Ballesteros, F. J. López, and M. Mittelbach. 2004. Optimization of alkali-catalyzed transesterification of Brassica Carinata oil for biodiesel production. *Energy and Fuels* 18 (1):77-83.

Freedman, B., E. H. Pryde, and T. L. Mounts. 1984. Variables affecting the yields of fatty esters from transesterified vegetable oils. *Journal of the American Oil Chemists' Society* 61 (10):1638-1643.

Gaurav, Kumar, Richa Srivastava, and Ram Singh. 2012. Exploring Biodiesel: Chemistry, Biochemistry, and Microalgal Source. *International Journal of Green Energy* 10 (8):775-796.

Goldberg, David E., K. Deb, H. Kargupta, and G. Harik. 1993. Rapid, accurate optimization of difficult problems using fast messy genetic algorithms. Paper read at Proceedings of the Fifth International Conference on Genetic Algorithms.

Hoekman, S. Kent, Amber Broch, Curtis Robbins, Eric Ceniceros, and Mani Natarajan. 2011. Review of biodiesel composition, properties, and specifications. *Renewable and Sustainable Energy Reviews* (0).

Hsu, C.W., C.C. Chang, and C.J. Lin. 2003. A practical guide to support vector classification. *URL www. csie. ntu. edu. tw/~ cjlin/papers/guide/guide. pdf*.

Huang, Cheng-Lung, and Chieh-Jen Wang. 2006. A GA-based feature selection and parameters optimizationfor support vector machines. *Expert Systems with Applications* 31 (2):231-240.

Kalogirou, Soteris A. 2001. Artificial neural networks in renewable energy systems applications: a review. *Renewable and Sustainable Energy Reviews* 5 (4):373-401.

Kohavi, Ron. 1995. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. Paper read at Proceedings of the 14th international joint conference on Artificial intelligence, at Montreal, Canada.

Noam, Lior. 2010. Energy resources and use: The present (2008) situation and possible sustainable paths to the future. *Energy* 35 (6):2631-2638.

Ozgul-Yucel, S., and S. Turkay. 2002. Variables affecting the yields of methyl esters derived from in situ esterification of rice bran oil. *Journal of the American Oil Chemists' Society* 79 (6):611-614.

Puig-Arnavat, Maria, Joan Carles Bruno, and Alberto Coronas. 2010. Review and analysis of biomass gasification models. *Renewable and Sustainable Energy Reviews* 14 (9):2841-2851.

Rajendra, Machavaram, Prakash Chandra Jena, and Hifjur Raheman. 2009. Prediction of optimized pretreatment process parameters for biodiesel production using ANN and GA. *Fuel* 88 (5):868-875.

Ramadhas, A. S., S. Jayaraj, C. Muraleedharan, and K. Padmakumari. 2006. Artificial neural networks used for the prediction of the cetane number of biodiesel. *Renewable Energy* 31 (15):2524-2533.

Shiu, Pei-Jing, Setiyo Gunawan, Wen-Hao Hsieh, Novy S. Kasim, and Yi-Hsu Ju. 2010. Biodiesel production from rice bran by a two-step in-situ process. *Bioresource Technology* 101 (3):984-989.

Vapnik, Vladimir Naumovich. 1995. *The nature of statistical learning theory*. New York: Springer.

Wakil, M. A., Abul Kalam, H. H. Masjuki, and I. M. Rizwanul Fattah. 2014. Rice Bran: A Prospective Resource for Biodiesel Production in Bangladesh. *International Journal of Green Energy*.

Yuste, Antonio J., and M. Pilar Dorado. 2005. A Neural Network Approach to Simulate Biodiesel Production from Waste Olive Oil. *Energy and Fuels* 20 (1):399-402.

Yustianingsih, L., S. Zullaikah, and Y. H. Ju. 2009. Ultrasound assisted in situ production of biodiesel from rice bran. *Journal of the Energy Institute* 82 (3):133-137.

Fig. 1. The procedure of GA-ESIM

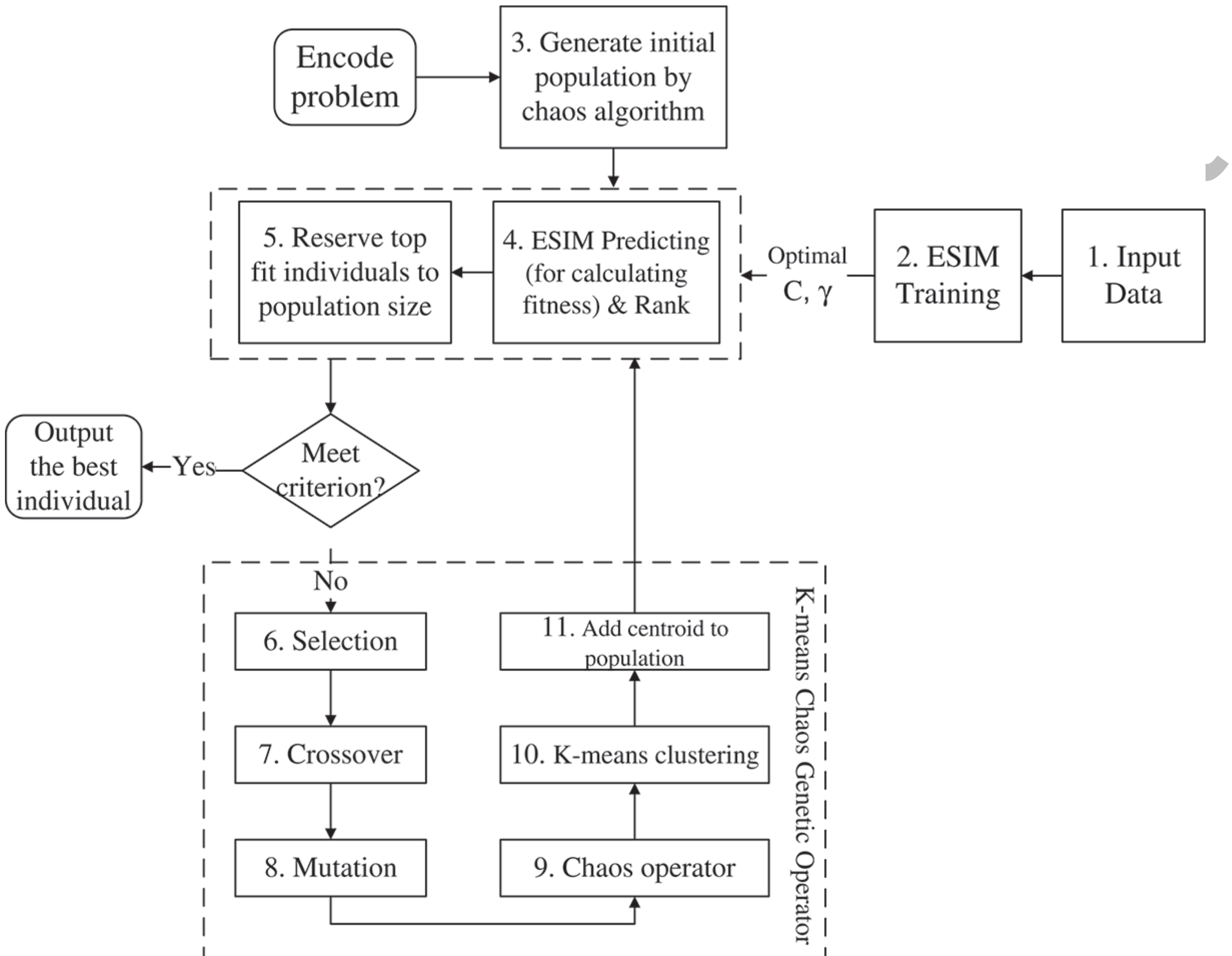

Fig. 2. ESIM Training and Testing result of fold 3

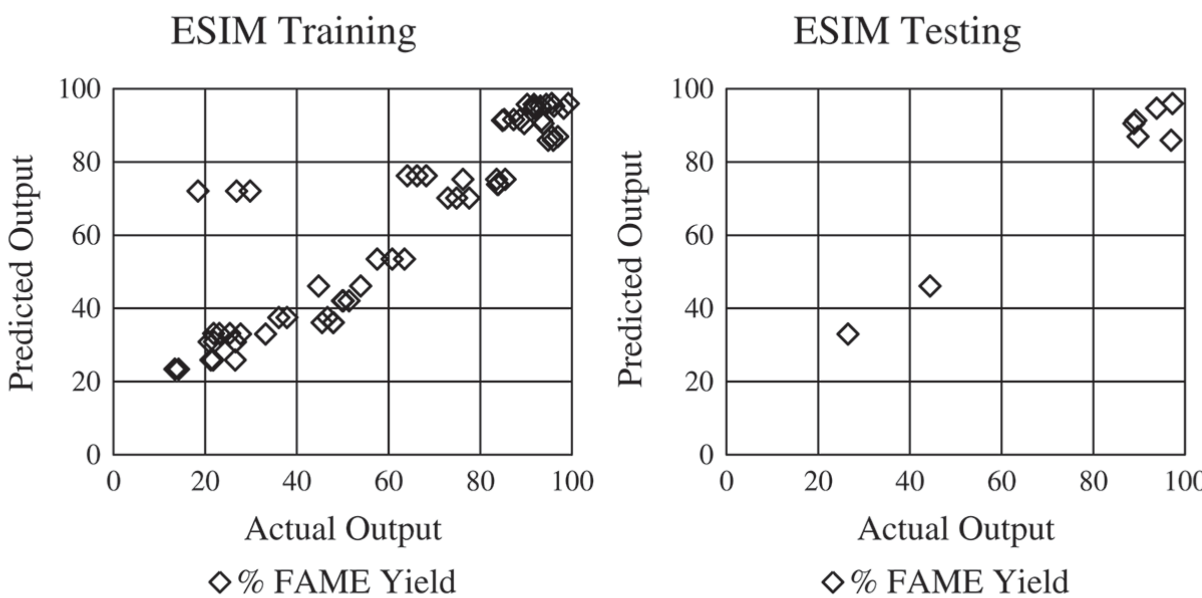

Fig. 3. GA-ESIM convergence graphics

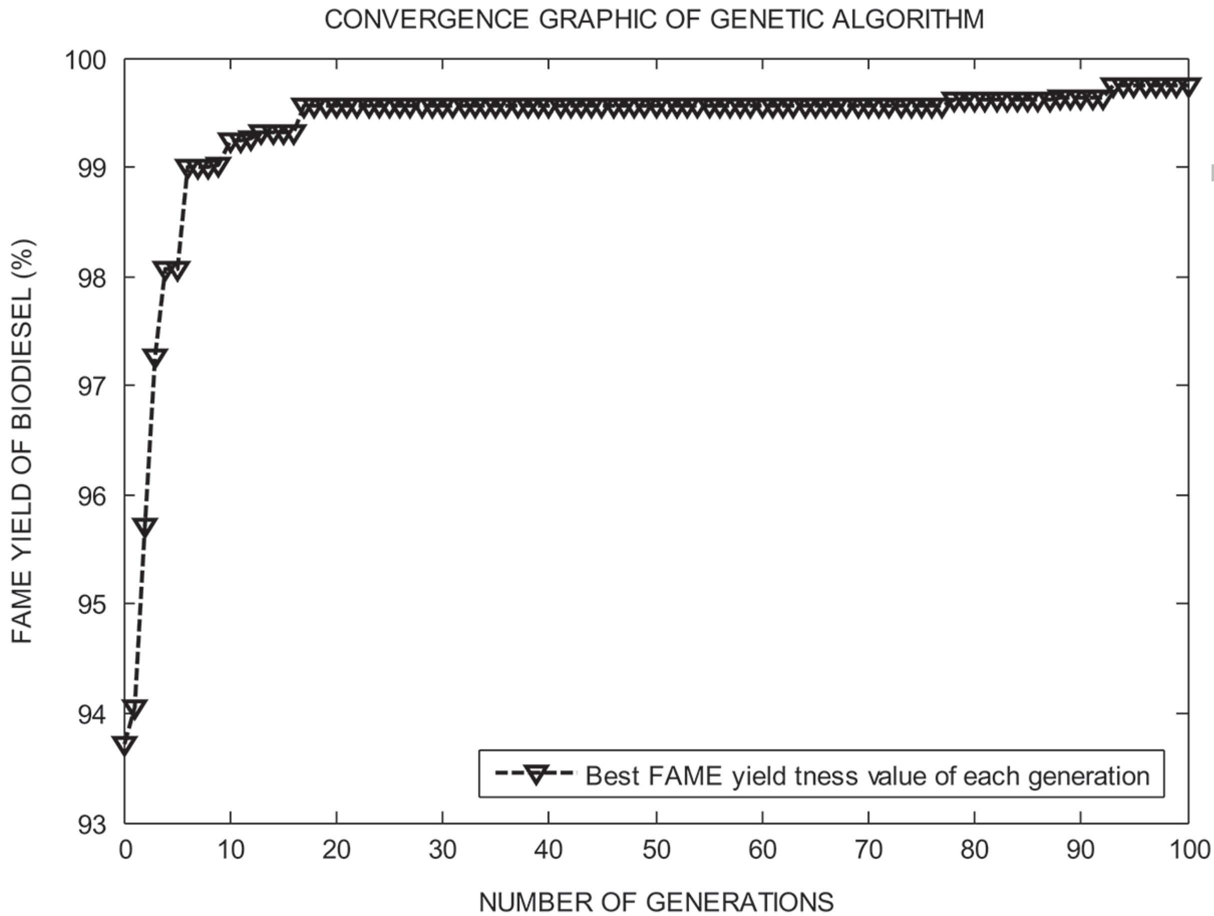

Fig. 4. Demonstration of Graphic User Interface (GUI) integrated with GA-ESIM

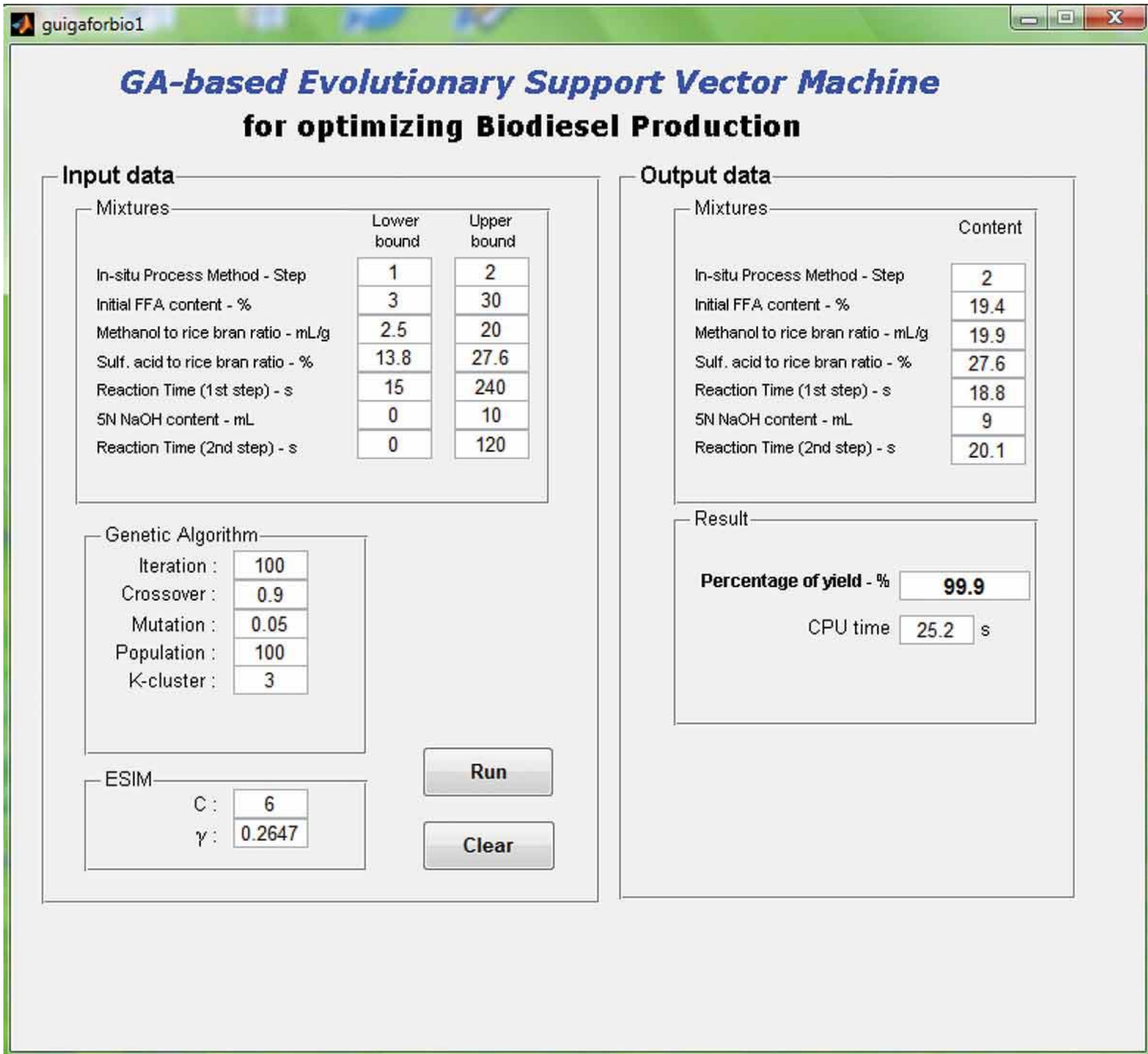

Table 1. Biodiesel production dataset

Case Number	In situ Process method (step)	Initial FFA content (%)	methanol to rice bran ratio (mL/g)	Sulfuric acid to rice bran mass ratio (%)	first step reaction time (minutes)	5N NaOH content (mL)	second step reaction time (minutes)	FAME yield (%)
1	1	3	2.5	27.6	60	0	0	20.85
2	1	3	2.5	27.6	60	0	0	26.61
3	1	3	2.5	27.6	60	0	0	22
4	1	3	5	27.6	60	0	0	27.71
5	1	3	5	27.6	60	0	0	26.5
6	1	3	5	27.6	60	0	0	33.18

7	1	3	10	27.6	60	0	0	36.05
8	1	3	10	27.6	60	0	0	37.83
9	1	3	10	27.6	60	0	0	46.66
10	1	3	15	27.6	60	0	0	49.95
11	1	3	15	27.6	60	0	0	51.36
12	1	3	15	27.6	60	0	0	50.13
13	1	3	20	27.6	60	0	0	44.41
14	1	3	20	27.6	60	0	0	53.93
15	1	3	20	27.6	60	0	0	44.75
16	1	3	15	27.6	15	0	0	21.85
17	1	3	15	27.6	15	0	0	25.36

18	1	3	15	27.6	15	0	0	23.1
19	1	3	15	27.6	30	0	0	47.92
20	1	3	15	27.6	30	0	0	47.96
21	1	3	15	27.6	30	0	0	45.5
22	1	3	15	27.6	60	0	0	49.95
23	1	3	15	27.6	60	0	0	51.36
24	1	3	15	27.6	60	0	0	50.13
25	1	3	15	27.6	120	0	0	63.48
26	1	3	15	27.6	120	0	0	60.78
27	1	3	15	27.6	120	0	0	57.5
28	1	3	15	27.6	240	0	0	72.89

29	1	3	15	27.6	240	0	0	77.55
30	1	3	15	27.6	240	0	0	74.84
31	1	3	15	13.8	15	0	0	13.39
32	1	3	15	13.8	15	0	0	14.18
33	1	3	15	18.4	15	0	0	21.73
34	1	3	15	18.4	15	0	0	21.24
35	1	3	15	18.4	15	0	0	26.53
36	1	3	15	27.6	15	0	0	21.85
37	1	3	15	27.6	15	0	0	25.36
38	1	3	15	27.6	15	0	0	23.1
39	2	3	15	27.6	15	7	120	29.81

40	2	3	15	27.6	15	7	120	18.44
41	2	3	15	27.6	15	7	120	26.85
42	2	3	15	27.6	15	8	120	83.93
43	2	3	15	27.6	15	8	120	83.74
44	2	3	15	27.6	15	8	120	83.89
45	2	3	15	27.6	15	9	120	76.19
46	2	3	15	27.6	15	9	120	85.4
47	2	3	15	27.6	15	9	120	83.55
48	2	3	15	27.6	15	10	120	64.06
49	2	3	15	27.6	15	10	120	68.22
50	2	3	15	27.6	15	10	120	66.2

51	2	3	15	27.6	15	8	5	88.6
52	2	3	15	27.6	15	8	5	87.26
53	2	3	15	27.6	15	8	5	85.22
54	2	3	15	27.6	15	8	15	84.83
55	2	3	15	27.6	15	8	15	93.39
56	2	3	15	27.6	15	8	15	89.28
57	2	3	15	27.6	15	8	30	89.59
58	2	3	15	27.6	15	8	30	93.7
59	2	3	15	27.6	15	8	30	88.85
60	2	3	15	27.6	15	8	60	96.88
61	2	3	15	27.6	15	8	60	89.76

62	2	3	15	27.6	15	8	60	95.6
63	2	3	15	27.6	15	8	120	83.93
64	2	3	15	27.6	15	8	120	83.74
65	2	3	15	27.6	15	8	120	83.89
66	2	30	15	27.6	15	8	5	96.07
67	2	30	15	27.6	15	8	5	93.12
68	2	30	15	27.6	15	8	15	90.21
69	2	30	15	27.6	15	8	15	91.7
70	2	30	15	27.6	15	8	15	94.4
71	2	30	15	27.6	15	8	30	97.27
72	2	30	15	27.6	15	8	30	95.57

73	2	30	15	27.6	15	8	30	99.21
74	2	30	15	27.6	15	8	60	93.84
75	2	30	15	27.6	15	8	60	91.74
76	2	30	15	27.6	15	8	60	98.17
77	2	30	15	27.6	15	8	120	96.98
78	2	30	15	27.6	15	8	120	94.81
79	2	30	15	27.6	15	8	120	95.9

Table 2. Biodiesel production influencing factors data

Input Factor	Unit	Upper bound	Lower bound
In-situ Process method	(Step)	1	2
Initial Free Fatty Acid (FFA) content	(%)	3	30
Methanol to rice bran ratio	(mL/g)	2.5	20
Sulfuric acid to rice bran ratio	(%)	13.8	27.6
Reaction time (first step)	(minutes)	15	240
5N NaOH content	(mL)	0	10
Reaction time (second step)	(minutes)	0	120
Output Factor	**Unit**	**Upper bound**	**Lower bound**

FAME yield	%	13.4	97.4

Table 3. Training and testing result of various prediction tools verified by 10-fold cross validation

Number of fold	Training RMSE (% FAME yield)		Testing RMSE (% FAME yield)		ESIM optimal parameters	
	ESIM	SVM	ESIM	SVM	C	γ
1	11.43	12.58	6.00	8.03	105	0.1579
2	10.66	12.67	5.62	9.55	198	0.2122
3	12.20	12.74	4.80	6.22	6	0.2647
4	11.41	12.64	6.34	9.37	16	0.5422
5	7.42	12.55	6.53	9.95	200	0.6401
6	6.98	9.71	8.55	25.31	195	0.9985
7	10.78	12.60	5.26	7.21	200	0.2001

8	10.03	12.52	7.54	9.97	190	0.3073
9	7.08	11.46	7.64	17.57	200	0.9751
10	7.45	12.47	6.63	8.36	110	0.9206
Minimum	6.98	9.71	4.80	6.22		
Average	9.54	12.20	6.49	11.16		
Maximum	12.20	12.74	8.55	25.31		

Table 4. ESIM training result of fold 3

Case Number	Desired Output (% FAME yield)	Estimated Output	
		ESIM (% FAME yield)	SVM (% FAME yield)
17	25.36	33.10	35.36
12	50.13	42.06	41.89
32	14.18	23.35	23.39
7	36.05	37.50	37.28
70	94.40	95.68	93.67
2	26.61	30.85	30.93
78	94.81	85.90	85.90
8	37.83	37.50	37.28

…	…	…	…
28	72.89	70.11	67.55
66	96.07	95.21	93.80
26	60.78	53.45	50.84
58	93.70	90.44	83.73
53	85.22	91.54	85.13
12	50.13	42.06	41.89
54	84.83	91.31	84.65
C, γ		6, 0.2647	1, 0.1429
RMSE		12.20	12.74

Table 5. ESIM testing result of fold 3

Case Number		Estimated Output	
	Desired Output (% FAME yield)	ESIM (% FAME yield)	SVM (% FAME yield)
77	96.98	85.90	85.90
59	88.85	90.44	83.73
61	89.76	86.88	81.20
56	89.28	91.31	84.65
13	44.41	46.09	46.46
5	26.50	32.94	32.92
71	97.27	95.91	93.27
74	93.84	94.65	91.71

C, γ	6, 0.2647	1, 0.1429
RMSE	4.80	6.22

Table 6. Mixture properties comparison between GA-ESIM result and the highest FAME yield record

Attribute	Unit	GA-ESIM	Highest FAME Yield Record
In-situ Process method	(Step)	2	2
Initial Free Fatty Acid (FFA) content	(%)	19.4	30
Methanol to rice bran ratio	(mL/g)	19.9	15
Sulfuric acid to rice bran ratio	(%)	27.6	27.6
Reaction time (first step)	(minutes)	18.8	15
5N NaOH content	(mL)	9	8
Reaction time (second step)	(minutes)	20.1	30
FAME yield	%	99.9	97.4